Research Paper | Education | E-ISSN No : 2454-9916 | Volume : 9 | Issue : 8 | August 2023

A COMPARISON OF THE UTILITY OF VARIOUS NEURAL NETWORK MODELS IN IMPROVING EDUCATION AND DESIGNING LEARNING PATHS

Aashaya Anand

Research Scholars Program, Harvard Student Agencies, In Collaboration with Learn with Leaders

ABSTRACT

Neural networks are self-improving computational systems used for prediction. Artificial Neural Networks (ANNs) computationally process information in a way that is similar to the human brain. There are myriad existing prediction models that can be used for various purposes, and this report aims to identify the predictive model most useful in the realm of education. It is taken into account that different students require different types of media to learn most effectively. In this project, different predictive models are compared to one another in their effectiveness specifically in predicting learning performance in certain subjects. Additionally, various activations (i.e., tanh, sigmoid, identity) and filtering methods (i.e., content-based, collaborative, and hybrid) are compared. These findings are then used to describe a possible recommendation algorithm to improve education by creating learning paths.

KEYWORDS: Neural Network, Recommendation Algorithm, ANN, Machine Learning

INTRODUCTION

Artificial Neural Networks (ANNs), analogous to biological neural networks, are non-linear models at the heart of machine learning. They are multilayered models intended to make calculated decisions like those of a human brain. Different predictive algorithms serve different purposes, and ANNs have proven to be quite effective for recommendation. Inputs entered in a neural network may be miscellaneous, but they are varied in importance by weights (values that enlarge an input relative to others, rendering them more important). A transfer function turns an input signal into an output signal. It is governed by a bias, which is what determines whether or not an item moves on to the next layer. An activation function is a nonlinear transformation applied to an item in order to present it in a way dictated by parameters. These processes once tried and repeated, produce a singular output, which in this case, is a recommendation. This is similar to recommendation algorithms used in platforms such as Pinterest and Spotify and can translate quite seamlessly into education resources.

Recommendation algorithms also involve filtering that is either *content-based, collaborative* or *hybrid.* These are based on similarities either between users or between items. The amount of similarity is determined by a dot product. This way, with a network of users and a database of content, an effective recommendation system can be created for each user. All of the previously mentioned terms and processes are further elaborated on in the literature review.

LITERATURE REVIEW

A conventional ANN consists of a Multilayer Perceptron, nodes at every layer, and nonlinear activation functions. A multilayer perceptron is a system of input and output layers, with a number of hidden layers. Figure 1 shows a rudimentary structure of a neural network.

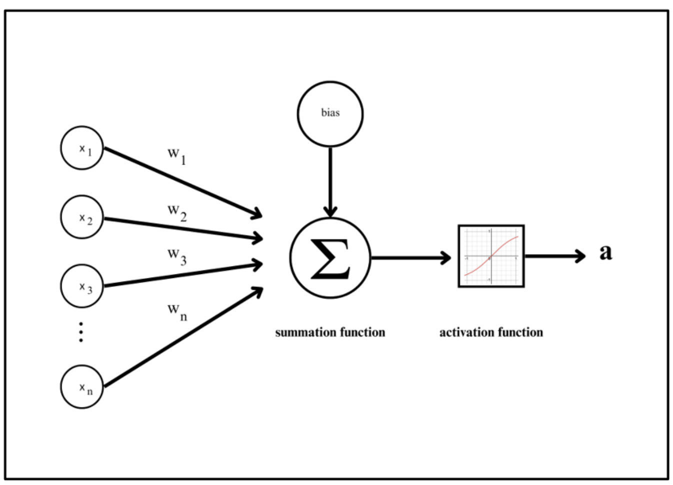


Figure 1: Basic Neural Network Structure

In the figure, x_1, x_2, x_3 … xn refer to input items, and w_1, w_2, w_3 … wn refer to their corresponding weights. Weights are numerical parameters multiplied by the input to determine the relative strength of neurons. They are meant to transform the input [1]. A transfer function (i.e., the point at which all weighted inputs converge) converts input signals into output signals. Here, biases are important in shifting the input either to the left or to the right. An activation function is used to determine a consolidated output. Depending on the function, the output is presented within certain parameters.

$$a = activation(w_i \times x_i + b);\ i \in \mathrm{N};\ b = bias$$

Activation functions are conventionally of two types: linear and nonlinear. Figure 2 shows the linear activation function or the identity function.

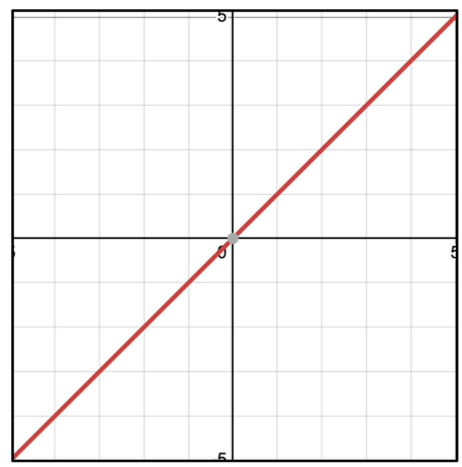


Figure 2: y = x or the Identity Function
Source: Desmos Online Graphing Calculator.

The caveat with using a linear activation function is that a neural network with a linear activation function and a number of layers is fundamentally the same as a neural network with no hidden layers [2]. This principle is based on the concept of back propagation.

After forward propagation, a loss function (that is meant to compare the target output and the real output) is produced. In order to minimise the 'loss' or discrepancy, weights need to be adjusted in the first stage of the network. This is called back propagation, and it ensures the algorithm is self-improving. This isn't possible in the presence of a linear activation function.

Different activation functions can be compared in terms of their domains and ranges [3].

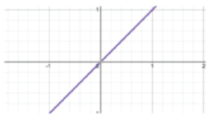
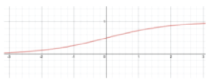
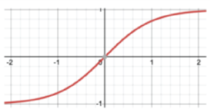
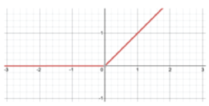
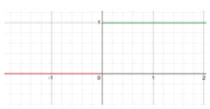
Activation Function	Graph	Equation
Identity Function		y = x
Sigmoid Function		$\phi(x) = \frac{1}{1+e^{-x}}$
Tanh		$tanh(x) = \frac{2}{1+e^{-2x}} - 1$
ReLU (Rectified Linear Unit)		$ReLU(x) = max(0, x)$
Binary Step Function		$f(x) = \begin{cases} y = 0; & x < 0 \\ y = 1; & x \geq 0 \end{cases}$

Table 1: Various Activation Functions
Source: Desmos Online Graphing Calculator [4]

Although the tanh and sigmoid functions appear to be similar, they vary in their range. The range of the sigmoid function is between 0 and 1, while the range of the tanh function is between -1 and 1, allowing it a lot more flexibility [5].

Learning path design must function on the principle of recommendation algorithms, to propel learning forward. Learning recommendation systems can be of three types: Collaborative Filtering, Content-based filtering, and Hybrid Filtering. Recommendation systems are based on rating each input in its validity to the situation in question. Outputs are ranked, and the most fit outputs are returned to the user. They involve Candidate Generation (choosing inputs from a plethora of options), Scoring (ordering), and Re-ranking (repeating after accounting for larger considerations) [6].

Content-based Filtering works on the principle of similarity between items; if items A and B are similar, a person who enjoyed item A will be recommended item B. A dot product is used as a measure of overlap or similarity.

The dot product of 2 matrices:

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix} \cdot \begin{bmatrix} w & x \\ y & z \end{bmatrix} = \begin{bmatrix} aw + by & ax + bz \\ cw + dy & cx + dz \end{bmatrix}$$

Collaborative Filtering works on the principle of similarity between users. If person 1 enjoyed item A and person 2 enjoyed item B, and if person 1 and 2 are similar, they will have similar recommendations. However, hybrid filtering (which is an amalgamation of the previous two methods) is the most preferred because of its greater accuracy.

Methods & Materials

The methodology of this project involves two parts: (I) using secondary data and different combinations of predictive models & activation functions to establish the most efficient combination and (II) using previous findings to describe a possible algorithm for a model that predicts the most efficient learning path.

I. For the first part of the study, various secondary sources were obtained from previous studies in which different neural network algorithms were developed for the purpose of predicting students' test scores. All data sets used were of university students pursuing degrees in STEM and CS. All data sets chosen were of a specific kind, so it is worth noting that the effectiveness of neural networks in predicting scores will not translate into humanities subjects in the same manner. This is because of the subjective nature of examinations and study material in subjects such as History, Music, Philosophy, and so forth. In all studies and sources, the equation for average prediction accuracy was as follows:

Wherein P_i refers to the number of correct predictions and Ai refers to the total number of predictions.

The algorithms compared were (1) Multilayer Perceptron Neural Network using the linear sigmoid activation function; (2) Multiple Linear Regression; (3) Radial Basis Function Network; (4) Support Vector Machine; (5) Deep Neural Network; and (6) a Feed Forward Artificial Neural Network using the tanh activation function [7][8][9][10][11].

II. The second part of this study involved using findings from part I to design a potential algorithm that creates an optimal learning path for a student. The same algorithms would work for this purpose too, because the algorithm must prove to account for niche improvements in performance and better performance in certain facets of the curriculum than others. A recommender system with an ANN at the heart of it is used.

A hybrid filtering process, involving content-based and collaborative filtering, is used to produce a recommendation by the criteria of duration of learning material, medium of learning material, areas of improvement, and so forth.

The content-based filtering stage would involve students establishing preferences in terms of media of education.

Tried	Enjoyed	Recommended
Video & Audio	Video	Video-based education
Text & Images	Text	Written resources
Physical & Virtual	Physical	Hands-on resources

Table 2: An example of a Content-based Filtering Stage

The collaborative filtering stage would compare the interests of various users to establish users that are similar. Then, the algorithm would cross-recommend similar items.

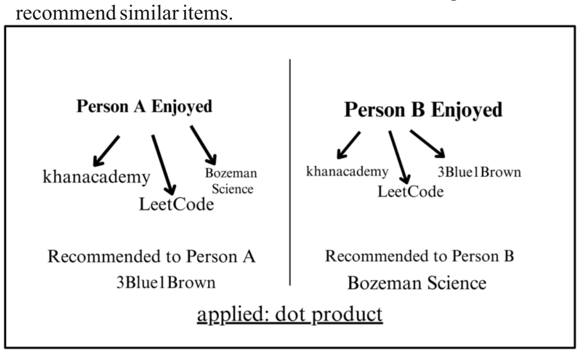


Figure 3: An example of a Collaborative Filtering Stage

Then, selected items would be run through an ANN to produce a recommendation to the user.

Through zero initialisation (setting all weights as 0), the first epoch will be carried out. After a loss function is generated, the process is repeated with random values. Based on the interest of the user, weights for each input are assigned. After all layers, biases are added to bridge gaps and errors.

Result 1

The following is the result of a study to compare the effectiveness of various predictive algorithms in forecasting the scores of college students through 1-2 semesters of university.

Serial Number	Model	Demographic	Sample Size (students)	Average Prediction Accuracy (APA)
1	**Multilayer perceptron neural network using the linear sigmoid activation function**	BS Engineering	150	84.5%

2	**Multiple Linear Regression (MLR)**	BS Engineering Dynamics	323	88.6%
3	**Radial Basis Function (RBF) Network**	BS Engineering Dynamics	323	88.3%
4	**Support Vector Machine (SVM)**	BS Engineering Dynamics	323	87.9%
5	**Deep Neural Network (DNN) + Transfer Learning**	BS Chemistry	282	77.68%
6	**ANN - Feed Forward & linear tanh**	University CS	4541	98.41%

Table 3: Relative Effectiveness of Various Predictive Algorithms in Predicting Test Scores

Discussion 1
It is evident from the data in Table 2 that an ANN with a feed-forward algorithm using the tanh function as its activation function, is the most effective algorithm for the purpose of predicting scores.

A possible reason why the 'multilayer perceptron neural network using the linear sigmoid activation function' model did not prove to be as accurate is that the tanh function is almost always preferred over the sigmoid function for such predictions.

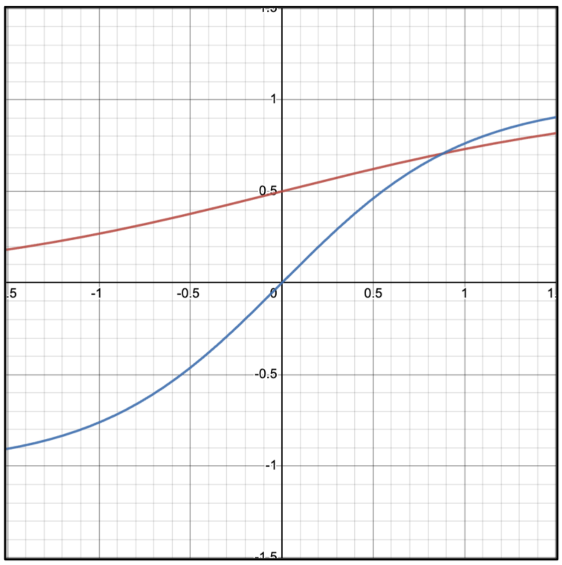


Figure 4: Sigmoid versus tanh Functions
Source: Desmos Online Graphing Calculator. Blue = tanh; red = sigmoid.

While both functions are nonlinear, the sigmoid function tends to push the output to one of the two extremes of the range (i.e., 0 and 1). This means that for values above 5, the output tends to be nearly 1, whereas, for values below -5, the output tends to be nearly 0. Also, since the range of the sigmoid function is between 0 and 1, it is best for predicting probabilities.

The hyperbolic tangent function can be interpreted as a stretched-out version of the sigmoid function. The key difference is that the tanh function is centred around 0, which is also true for most datasets. This results in a normal-like distribution of outputs.

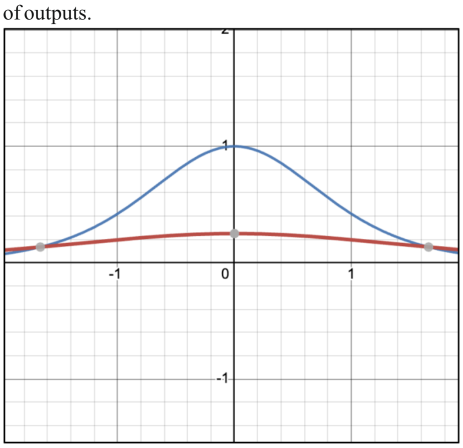


Figure 5: Derivatives of Sigmoid versus tanh Functions
Source: Desmos Online Graphing Calculator. Blue = tanh; red = sigmoid.

Support Vector Machine (SVM) is a supervised learning algorithm most commonly used in the discipline of classified machine learning. This is likely a reason it was slightly ineffective in predicting test scores. It functions by creating decision boundaries to segregate data into classes. It requires a large number of epochs and a long training duration to function optimally and is also less effective for larger datasets. These are also reasons it may have been less effective [12].

Additionally, many deep neural networks or convolutional neural networks are more useful for the purpose of image, text, or character recognition than the prediction of test scores, prediction of stocks, weather reports, etc. This doesn't insinuate that they're less useful as neural networks in general; just that they are relatively less effective in this area specifically. This also goes for most other prediction algorithms and models assessed in this study. Other factors that may influence results are sample size (which was much larger for ANNs than any other category), amount of training time, number of epochs, and so forth. Also, the specific curriculum design was likely very different for each set of students assessed.

Result 2

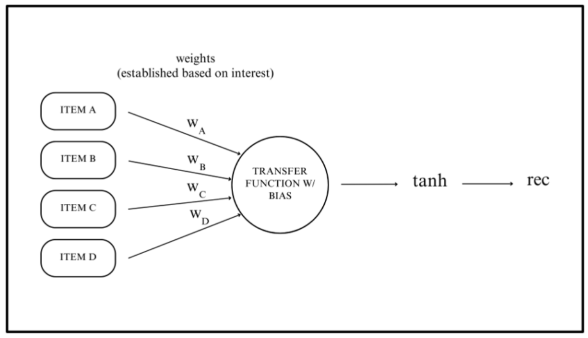


Figure 6: ANN based on Recommendation Algorithm to Produce a Recommendation for Learners

Using collaborative and content-based filtering, the interests and preferences of a user are found. These are reflected in the weights added to the inputs. For example, if a user enjoys visual information more than auditory information, a video explaining a concept will be weighed more than a voice recording. It will thus be more likely to be recommended to the user. After running all inputs through a transfer function and the tanh activation function, a recommendation is given to the user.

Discussion 2
This is a recommendation algorithm similar to those used in platforms such as Pinterest, Spotify, and so forth. A similar algorithm can be applied to educational content to optimise and enhance the experience of versatile learning.

CONCLUSION
Of all predictive algorithms, ANNs using the tanh activation function have proven to be the most effective in predicting test scores, as well as providing educational recommendations to users. The latter can also be supported by the fact that ANNs are used in recommendation algorithms in various other media platforms. Educational media and material being versatile and adequately stimulating is an important factor in learning. Since different users require different types of material in order to learn most effectively [13], content-based and collaborative filtering can be used to establish what is optimal for which user. Interest (whether explicitly or implicitly expressed) contributes to weights for each input in an ANN. The tanh function, which is a function centred around zero, is a great activation function for this purpose. Dot products can be used to find the amount of overlap between different types of users and content. All of the previous findings converge to an ANN that recommends the best type of educational media to a learner.

REFERENCES
1. Krogh, A. What are artificial neural networks? Nat Biotechnol 26, 195–197 (2008). https://doi.org/10.1038/nbt1386
2. Mohammadzaheri, M., Chen, L., Ghaffari, A., & Willison, J. (2009). A combination of linear and nonlinear activation functions in neural networks for modeling a de-superheater. Simulation Modelling Practice and Theory, 17(2), 398-407.
3. Karlik, B., & Olgac, A. V. (2011). Performance analysis of various activation functions in generalized MLP architectures of neural networks. International Journal of Artificial Intelligence and Expert Systems, 1(4), 111-122.
4. Desmos Graphing Calculator, https://www.desmos.com/calculator
5. Kalman, B. L., & Kwasny, S. C. (1992, June). Why tanh: choosing a sigmoidal function. In [Proceedings 1992] IJCNN International Joint Conference on Neural Networks (Vol. 4, pp. 578-581). IEEE.
6. Li, Q., & Kim, B. M. (2003, July). An approach for combining content-based and collaborative filters. In Proceedings of the sixth international workshop on Information Retrieval with Asian languages (pp. 17-24).
7. Samy Abu Naser, Ihab Zaqout, Mahmoud Abu Ghosh, Rasha Atallah, Eman Alajrami, Predicting Student Performance Using Artificial Neural Network: in the Faculty of Engineering and Information Technology, International Journal of Hybrid Information Technology Vol.8, No.2 (2015), pp.221-228

http://dx.doi.org/10.14257/ijhit.2015.8.2.20

8. Shaobo Huang, Ning Fang, Predicting student academic performance in an engineering dynamics course: A comparison of four types of predictive mathematical models, Computers & Education, Volume 61, 2013, Pages 133-145, ISSN 0360-1315, https://doi.org/10.1016/j.compedu.2012.08.015. (https://www.sciencedirect.com/science/article/pii/S0360131512002102)
9. Tsiakmaki, M., Kostopoulos, G., Kotsiantis, S., & Ragos, O. (2020). Transfer Learning from Deep Neural Networks for Predicting Student Performance. Applied Sciences, 10(6), 2145. MDPI AG. Retrieved from http://dx.doi.org/10.3390/app10062145
10. T. Wang and A. Mitrovic, "Using neural networks to predict student's performance," International Conference on Computers in Education, 2002. Proceedings., Auckland, New Zealand, 2002, pp. 969-973 vol.2, doi: 10.1109/CIE.2002.1186127.
11. Yağcı, Ali & Çevik, Mustafa. (2017). Predictions of academic achievements of vocational and technical high school students with artificial neural networks in science courses (physics, chemistry and biology) in Turkey and measures to be taken for their failures. SHS Web of Conferences. 37. 01057. 10.1051/shsconf/20173701057.
12. Suthaharan, S., & Suthaharan, S. (2016). Support vector machine. Machine learning models and algorithms for big data classification: thinking with examples for effective learning, 207-235.
13. Samuelowicz, K. (1987). Learning problems of overseas students: Two sides of a story. Higher Education Research and Development, 6(2), 121-133.